Reduced basis method for the rapid and reliable solution of partial differential equations

Yvon Maday

Reduced Basis Method for the Rapid and Reliable Solution of Partial Differential Equations

Yvon Maday*

Abstract. Numerical approximation of the solution of partial differential equations plays an important role in many areas such as engineering, mechanics, physics, chemistry, biology... for computer-aided design-analysis, computer-aided decision-making or simply better understanding. The fidelity of the simulations with respect to reality is achieved through the combined efforts to derive: (i) better models, (ii) faster numerical algorithm, (iii) more accurate discretization methods and (iv) improved large scale computing resources. In many situations, including optimization and control, the *same* model, depending on a parameter that is changing, has to be simulated over and over, multiplying by a large factor (up to 100 or 1000) the solution procedure cost of one simulation. The reduced basis method allows to define a surrogate solution procedure, that, thanks to the complementary design of fidelity certificates on outputs, allows to speed up the computations by two to three orders of magnitude while maintaining a sufficient accuracy. We present here the basics of this approach for linear and non linear elliptic and parabolic PDE's.

Mathematics Subject Classification (2000). 65D05, 65M60, 65N15, 65N30, 65N35.

Keywords. Reduced-basis, a posteriori error estimation, output bounds, offline-online procedures, Galerkin approximation, parametrized partial differential equations.

1. Introduction

Let us consider a class of problems depending on a parameter $\mu \in \mathcal{D}$ set in the form: find $u \equiv u(\mu) \in X$ such that $\mathcal{F}(u;\mu) = 0$ (we do not specify much at this point what $\mathcal{D}$ is, it could be a subset of $\mathbf{R}$, or $\mathbf{R}^p$, or even a subset of functions). Such problems arise in many situations such as e.g. optimization, control or parameter-identification problems, response surface or sensibility analysis. In case

*This paper presents a review of results (on the definition, analysis and solution strategies) most of them first presented elsewhere and that have benefitted from the long-standing collaboration with Anthony T. Patera, Einar M. Rønquist, Gabriel Turinici, and more recently Annalisa Buffa, Emil Løvgren, Ngoc Cuong Nguyen, Georges Pau, Christophe Prud'homme. This paper is not intended to be exhaustive and is an invitation to read more complete papers that treat in depth the features presented here. Due to this, no figures nor tables synthesizing numerical results are provided.

$\mathcal{F}$ is written through partial differential equations, the problem may be stationary or time dependent but in all these cases, a solution $u(\mu)$ has to be evaluated or computed for many instances of $\mu \in \mathcal{D}$. Even well optimized, the favorite discretization method of yours will lead to very heavy computations in order to approximate all these solutions and decision may not be taken appropriately due to too large computer time for reliable simulations.

The approach discussed in this paper will not aim at presenting an alternative to your favorite discretization, more the contrary. The idea is that, in many cases, your discretization will help in constructing a surrogate method that will allow to mimick it or at least to do the spadework on the evaluation of the optimal or control solution. The complexity of the equations resulting from this approach will be very low, enabling very fast solution algorithms. No miracle though, the method is based on a learning strategy concept, and, for a new problem, the preliminary *off-line* preparation is much time consuming. It is only after this learning step that the full speed of the method can be appreciated *on-line*, paying off the cost of the *off-line* preparation step. During the first step, we evaluate accurately, based on your prefered solver, a few solutions to $\mathcal{F}(u;\mu) = 0$; actually, any discretization method is good enough here. In the second step, that is involved on request and *on-line*, the discretization method that has been used earlier is somehow forgotten and a new discretization approach is constructed based on a new ad-hoc basis set (named "reduced basis") built out from the previous computations. In many cases the method proves very efficient and — even though the complete understanding of the reasons why it is working so well are not mastered — an a *posteriori* error theory allows to provide fidelity certificates on outputs computed from the reduced-basis -discretized solution. This method is valid in case the set $\mathcal{S}(\mathcal{D}) = \{u(\mu), \mu \in \mathcal{D}\}$ has a simple (hidden) structure, the solution $u(\mu)$ has to be regular enough in μ. We provide some explanations on the rational of the reduced basis approximation in section 2 and present the method in the elliptic case. In section 3 we give more rigorous explanation on the rapid convergence of the method on a particular case. This is complemented in section 4 by an analysis of a posteriori tools that provide fidelity certificate for outputs computed from the reduced basis approximation. Section 5 tells more about the track to follow to be convinced that the method will "work" on the particular problem of yours. The efficient implementation of the reduced basis method needs some care, we present in section 6 some of the required tools. Finally we end this paper by providing in section 7 some of the new directions we are currently working on.

2. Basics and rational of the reduced basis approach

The reduced basis method consists in approximating the solution $u(\mu)$ of a parameter dependent problem $\mathcal{F}(u;\mu) = 0$ by a linear combination of appropriate, preliminary computed, solutions $u(\mu_i)$ for well chosen parameters μ_i, $i = 1, \ldots, N$. The rational of this approach, stands in the fact that the set $\mathcal{S}(\mathcal{D}) = \{u(\mu)$ of all solutions when $\mu \in \mu\}$ behaves well. In order to apprehend in which sense the

good behavior of $\mathcal{S}(\mathcal{D})$ should be understood, it is helpfull to introduce the notion of n-width following Kolmogorov [8] (see also [14])

Definition 2.1. Let X be a normed linear space, A be a subset of X and X_n be a generic n-dimensional subspace of X. The deviation of A from X_n is

$$E(A; X_n) = \sup_{x \in A} \inf_{y \in X_n} \|x - y\|_X .$$

The *Kolmogorov n-width* of A in X is given by

$$\begin{aligned} d_n(A, X) &= \inf\{E(A; X_n) : X_n \text{ an } n\text{-dimensional subspace of X}\} \\ &= \inf_{X_n} \sup_{x \in A} \inf_{y \in X_n} \|x - y\|_X \ . \end{aligned} \tag{1}$$

The n-width of A thus measures the extent to which A may be approximated by a n-dimensional subspace of X. There are many reasons why this n-width may go rapidly to zero as n goes to infinity. In our case, where $A = \mathcal{S}(\mathcal{D})$, we can refer to regularity of the solutions $u(\mu)$ with respect to the parameter μ, or even to analyticity. Indeed, an upper bound for the asymptotic rate at which the convergence to zero is achieved is provided by this example from Kolmogorov stating that $d_n(\tilde{B}_2^{(r)}; L^2) = \mathcal{O}(n^{-r})$ where $\tilde{B}_2^{(r)}$ is the unit ball in the Sobolev space of all 2π-periodic real valued, $(r-1)$-times differentiable functions whose $(r-1)$st derivative is absolutely continuous and whose rth derivative belongs to L^2. Actually, exponential convergence is achieved when analyticity exists in the parameter dependency. The knowledge of the rate of convergence is not sufficient: of theoretical interest is the determination of the (or at least one) optimal finite dimensional space X_n that realizes the infimum in d_n, provided it exists. For practical reasons, we want to restrict ourselves to finite dimensional spaces that are spanned by elements of $\mathcal{S}(\mathcal{D})$. This should increase the n-width, but it appears that it does not increase too much as is explained in the following result; based upon [14], we can prove

Theorem 2.2. *Let X be a Hilbert space and A be a bounded subset of X such that $d_n(A, X) \neq 0$. Let X_A denote the vectorial space spanned by A, then*

$$d_n(A, X_A) \leq (n+1)\ d_n(A, X) \ .$$

Proof. Since the width of A and of its symmetric, closed, convex hull are equal, we shall assume hereafter that A is closed, convex and symmetric. For any $z_1, \ldots, z_m \in X$ let us denote by $V_m(z_1, \ldots, z_m)$ the volume of the parallelipiped determined by the vectors $z_1, \ldots, z_m$. Let us define

$$V = \sup\{V_{n+1}(z_1, \ldots, z_{n+1}) : z_k \in A\}.$$

Since A is bounded, $V < \infty$. If $V = 0$, then A is contained in a n dimensional space which is in contradiction with $d_n(A, X) \neq 0$, then for $\varepsilon > 0$ sufficiently small, we can choose $x_1, \ldots, x_{n+1} \in A$ such that

$$V_{n+1}(x_1, \ldots, x_{n+1}) > (1 - \varepsilon)V > 0 \ .$$

Let us introduce the space

$$E_k = \text{Span}\{x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_{n+1}\} \ ,$$

then the following equality, for any $y_1 \in X$, is straightforward

$$V_{n+1}(y_1, x_2, \ldots, x_{n+1}) = E(y_1; E_1) V_n(x_2, \ldots, x_{n+1}) \ .$$

From the definition of $d_n(A, X_A)$, there exists $\tilde{x}_1 \in A$ for which $E(\tilde{x}_1; E_1) \geq d_n(A, X_A)$, hence:

$$\begin{aligned} V &\geq V_{n+1}(\tilde{x}_1, x_2, \ldots, x_{n+1}) \\ &= E(\tilde{x}_1; E_1) V_n(x_2, \ldots, x_{n+1}) \\ &= \frac{E(\tilde{x}_1; E_1)}{E(x_1; E_1)} V_{n+1}(x_1, \ldots, x_{n+1}) \\ &\geq \frac{d_n(A, X_A)}{E(x_1; E_1)} V_{n+1}(x_1, \ldots, x_{n+1}) \ . \end{aligned} \tag{2}$$

Therefore $E(x_1; E_1) \geq (1-\varepsilon) d_n(A, X_A)$, and the same is true for any k: $E(x_k; E_k) \geq (1-\varepsilon) d_n(A, X_A)$. By Lemma 5.4 and Proposition 1.6 in [14] we derive that $(1-\varepsilon) d_n(A, X_A) \leq (n+1) d_n(A, X)$. Since this holds for any $\varepsilon > 0$, the theorem follows. □

We derive from the theorem that the quantity

$$\inf\{ \sup_{u \in \mathcal{S}(\mathcal{D})} \inf_{y \in X_n} \|x - y\|_X : X_n = \text{Span}\{u(\mu_1), \ldots, u(\mu_n), \mu_i \in \mathcal{D}\}\} \tag{3}$$

converges to zero (almost at the same speed as $d_n(\mathcal{S}(\mathcal{D}); X)$ provided very little regularity exists in the parameter dependancy of the solution $u(\mu)$, and an exponential convergence is achieved in many cases since analyticity in the parameter is quite frequent.

This is at the basics of the reduced basis method. Indeed we are led to choose properly a sequence of parameters $\mu_1, \ldots, \mu_n, \cdots \in \mathcal{D}$, then define the vectorial space $X_N = \text{Span}\{u(\mu_1), \ldots, u(\mu_N)\}$ and look for an approximation of $u(\mu)$ in X_N.

Let us consider for example an elliptic problem : Find $u(\mu) \in X$ such that

$$a(u(\mu), v; \mu) = f(v), \quad \forall v \in X \ . \tag{4}$$

Here X is some Hilbert space, and a is a continuous and elliptic, bilinear form in its two first arguments, regular in the parameter dependancy and f is some given continuous linear form. We assume for the sake of simplicity that the ellipticity is uniform with respect to $\mu \in \mathcal{D}$: $\exists \alpha > 0$

$$\forall \mu \in \mathcal{D}, \forall u \in X, \quad a(u, u; \mu) \geq \alpha \|u\|_X^2 \ ,$$

an that the continuity of a is uniform with respect to $\mu \in \mathcal{D}$ as well: $\exists \gamma > 0$

$$\forall \mu \in \mathcal{D}, \forall u, v \in X, \quad |a(u, v; \mu)| \leq \gamma \|u\|_X \|v\|_X \ .$$

It is classical to state that, under the previous hypothesis, problem (4) has a unique solution for any $\mu \in \mathcal{D}$. The Galerkin method is a standard way to approximate the solution to (4) provided that a finite dimensional subspace X_N on X is given. It consists in : Find $u_N(\mu) \in X_N$ such that

$$a(u_N(\mu), v_N; \mu) = f(v_N), \quad \forall v_N \in X_N \quad , \tag{5}$$

which similarly has a unique solution $u_N(\mu)$. Cea's lemma then states that

$$\|u(\mu) - u_N(\mu)\|_X \leq (1 + \frac{\gamma}{\alpha}) \inf_{v_N \in X_N} \|u(\mu) - v_N\|_X \ . \tag{6}$$

The best choice for the basis element $u(\mu_1), \ldots, u(\mu_N)$ of X_N would be those that realize the infimum in (3), i.e. the ones that realize the maximum of the volume $V_N(u(\mu_1), \ldots, u(\mu_N))$. Unfortunately, this is not a constructive method and we generally refer to a greedy algorithm such as the following one:

$$\begin{aligned} \mu_1 &= \arg\sup_{\mu \in \mathcal{D}} \|u(\mu)\|_X \ , \\ \mu_{i+1} &= \arg\sup_{\mu \in D} \|u(\mu) - P_i u(\mu)\|_X \ , \end{aligned} \tag{7}$$

where P_i is the orthogonal projection onto $V_i = \text{span}\{u(\mu_1), \ldots, u(\mu_i)\}$ or a variant of it that is explained at the end of section 4. The convergence proof for the related algorithm is somehow more complex and presented in a quite general settings in [1].

3. An example of *a priori* analysis.

The previous notion of n-width is quite convenient because it is rather general, in spirit, and provides a tool to reflect the rapid convergence of the reduced basis method but it is not much constructive nor qualitatively informative. We are thus going to consider a particular example where the parametrized "bilinear" form $a\colon X \times X \times \mathcal{D} \to \mathbf{R}$ is defined as follows

$$a(w, v; \mu) \equiv a_0(w, v) + \mu a_1(w, v) \ ; \tag{8}$$

here the bilinear forms $a_0\colon X \times X \to \mathbf{R}$ and $a_1\colon X \times X \to \mathbf{R}$ are continuous, symmetric and positive semi-definite, $\mathcal{D} \equiv [0, \mu_{\max}]$, and we assume that a_0 is coercive. It follows from our assumptions that there exists a real positive constant γ_1 such that

$$0 \leq \frac{a_1(v, v)}{a_0(v, v)} \leq \gamma_1, \quad \forall\, v \in X \ . \tag{9}$$

For the hypotheses stated above, it is readily demonstrated that the problem (4) satisfies uniformly the Lax Milgram hypothesis.

Many situations may be modeled by this rather simple problem statement (4), (8). It can be the conduction in thin plates and μ represents the convective heat

transfer coefficient, it can also be a variable-property heat transfer, then $1+\mu$ is the ratio of thermal conductivities in domains ...

The analysis that we did in [12] involves the eigenproblem : Find $(\varphi \in X, \lambda \in \mathbf{R})$, satisfying $a_1(\varphi, v) = \lambda\, a_0(\varphi, v)$, $\forall\, v \in X$. Indeed the solution $u(\mu)$ to problem (4) can be expressed as

$$u(\cdot, \mu) = \int \frac{f(\varphi)\, \varphi(\cdot; \lambda)}{1+\mu\lambda} d\lambda \ . \tag{10}$$

The dependancy in μ is thus explicitly expressed and we can propose to approximate $u(\mu)$ by a linear combination of well chosen $u(\mu_i)$. This can be done through interpolation at the μ_i by polynomials. It is interesting to notice at this point that we have a large choice in the variable in which the polynomial can be expressed. Indeed since we are interested through this interpolation process to a evaluate the best fit, a polynomial in μ may not be the best choice but rather a polynomial in $\frac{1}{\mu}$, e^{μ} or else.... in [12] we have considered a polynomial approximation in the variable $\tau = \ln(\mu + \delta^{-1})$, where δ is some positive real number. The analysis then involves the interpolation operator at equidistant points (in the τ variable) for which we were able to get an upper bound used, in turn, to qualify the best fit result

Lemma 3.1. *There exists a constant $C > 0$ and a positive integer $N_{\rm crit}$ such that for $N \geq N_{\rm crit}$*

$$\inf_{w_N \in X_N} \|u(\mu) - w_N\|_X \leq C \exp\left\{\frac{-N}{N_{\rm crit}}\right\}, \qquad \forall\, \mu \in \mathcal{D} \ ,$$

where $N_{\rm crit} \equiv c^* e\, \ln(\gamma\, \mu_{\max} + 1)$.

This analysis of [12] leads to at least three remarks :

Remark 3.2. a) The analysis of the best fit done here suggests to use sample points μ_i that are equidistant when transformed in the τ variable. We performed some numerical tests to check whether this sampling gives indeed better results than more conventional ones (of course you should avoid equidistant in the original μ variable, but we tried e.g. Chebyshev points) and this was actually the case. Unfortunately, in more general situations and especially in higher parameter dimensions, we have no clue of a direct constructive best sampling method.
b) For a given sampling μ_i, one can propose an interpolation procedure to approximate $u(\mu)$ which is more simple than referring to a Galerkin approach. Indeed, an approximation

$$u(\mu) \simeq \sum_{i=1}^{N} \alpha_i(\mu) u(\mu_i) \ ,$$

can be proposed by using coefficients that are the Lagrange interpolation basis in the chosen variable (above it was $\tau = \ln(\mu+\delta^{-1})$, i.e. the mapping $\tau \mapsto \alpha_i(\mu(\tau))$ is a polynomial of degree $\leq N$ and $\alpha_i(\mu_j) = \delta_{ij}$). The problem is that the expression of $\alpha_i(\mu)$ depends on the best choice of variable which is unknown and within a set that is quite infinite providing a range of results that are quite different. Since

for a given general problem we have no clue of the best interpolation system, the Galerkin approach makes sense, indeed .....
c) In opposition, the Galerkin approach does not require any preliminary analysis on guessing the way the solution depends upon the parameter. Its superiority over interpolation process comes from the fact stated in Cea's lemma that the approximation that is obtained, up to some multiplicative constant, gives the optimal best fit, even if we do not know the rate at which the convergence is going.

Finally, as is often the case, we should indicate that the a priori analysis helps to have confidence in developing the method but, at the end of a given computation, a computable a posteriori estimator should be designed in order to qualify the approximation. This is even more true with such new surrogate approximation in order to replace the expertise a user may have in his preferred method, e.g. his intuition on the choice of the discretization parameter to get acceptable discrete solutions. This is the purpose of the next section.

4. An example of *a posteriori* analysis.

Most of the time, the complete knowledge of the solution of the problem (4) is not required. What is required, is outputs computed from the solution $s = s(u)$, where s is some continuous functional defined over X. In order to have a hand over this output, the reduced basis method consists first in computing $u_N \in X_N$ solution of the Galerkin approximation (5), then propose $s_N = s(u_N)$ as an approximation of s. Assuming Lipschitz condition (ex. linear case) over s, it follows that

$$|s - s_N| \leq c\|u - u_N\|_X. \tag{11}$$

Thus any information over the error in the energy norm will allow to get verification (provided you are able to evaluate c). Actually it is well known that the convergence of s_N towards s most often goes faster. This is standard but we go back over it since this will prove usefull in the sequel. Let us assume we are in the linear output case where $s \equiv \ell$ is a linear continuous mapping over X. It is then standard to introduce the $a$djoint state, solution of the following problem : find $\psi \in X$

$$a(v, \psi; \mu) = -\ell(v), \quad \forall v \in X. \tag{12}$$

The error in the output is then (remember that, for any $\phi_N \in X_N$, $a(u, \phi_N; \mu) = a(u_N, \phi_N; \mu) = (f, \phi_N)$)

$$\begin{aligned} s_N - s &= \ell(u_N) - \ell(u) \\ &= a(u, \psi; \mu) - a(u_N, \psi; \mu) \\ &= a(u, \psi - \phi_N; \mu) - a(u_N, \psi - \phi_N; \mu), \quad \forall \phi_N \in X_N \\ &= a(u - u_N, \psi - \phi_N; \mu), \quad \forall \phi_N \in X_N \\ &\leq c\|u - u_N\|_X \|\psi - \phi_N\|_X, \quad \forall \phi_N \in X_N \ , \end{aligned} \tag{13}$$

so that the best fit of ψ in X_N can be chosen in order to improve the first error bound (11) that was proposed for $|s - s_N|$.

For instance if ψ_N is the solution of the Galerkin approximation to ψ in X_N, we get

$$|s - s_N| \leq c\|u - u_N\|_X \|\psi - \psi_N\|_X \ . \tag{14}$$

Actually, the approximation of ψ in X_N may not be very accurate since X_N is well suited for approximating the elements $u(\mu)$ and — except in the case where $\ell = f$ named the complient case — a separate reduced space $\tilde{X}_N$ should be built which provides an associated approximation $\tilde{\psi}_N$. Then an improved approximation for $\ell(u)$ is given by $\ell_{\rm imp} = \ell(u_N) - a(u_N, \tilde{\psi}_N) + f(\tilde{\psi}_N)$ since (14) holds with $\|\psi - \tilde{\psi}_N\|_X$ for which a better convergence rate is generally observed.

Even improved, this result is still *a priori* business and it does not allow to qualify the approximation for a given computation. In order to get *a posteriori* information, between $\ell(u)$ and $\ell(u_N)$ (or $\ell_{\rm imp}$), we have to get a hand on the residuals in the approximations of the primal and dual problems. We introduce for any $v \in X$,

$$\mathcal{R}^{pr}(v;\mu) = a(u_N, v; \mu) - < f, v >, \quad \mathcal{R}^{du}(v;\mu) = -a(v, \tilde{\psi}_N; \mu) - \ell(v). \tag{15}$$

We then compute the reconstructed errors associated with the previous residuals. These are the solutions of the following problems

$$2\alpha \int \nabla \hat{e}^{pr(du)} \nabla v = \mathcal{R}^{pr(du)}(v;\mu), \quad \forall v \ , \tag{16}$$

we then have

Theorem 4.1. *Let $s^- = \ell_{\rm imp} - \alpha \int [\nabla(\hat{e}^{pr} + \hat{e}^{du})]^2$ then $s^- \leq s$. In addition, there exists two constants $0 < c \leq C$ such that*

$$c|s - s_N| \leq s - s^- \leq C|s - s_N|.$$

Proof. Let us denote by e_N the difference between the exact solution and the approximated one $e_N = u - u_N$. From (16), we observe that

$$2\alpha \int \nabla \hat{e}^{pr} \nabla e_N = -a(e_N, e_N; \mu)$$

and

$$2\alpha \int \nabla \hat{e}^{du} \nabla e_N = -a(e_N, \tilde{\psi}_N; \mu) - \ell(e_N) = f(\tilde{\psi}_N) - a(u_N, \tilde{\psi}_N) - \ell(e_N).$$

Taking this into account allows to write

$$\begin{aligned} \ell_{\rm imp} - \alpha \int \nabla(\hat{e}^{pr} + \hat{e}^{du})^2 &= \ell(u_N) - a(u_N, \tilde{\psi}_N) + f(\tilde{\psi}_N) - \alpha \int \nabla(\hat{e}^{pr} + \hat{e}^{du})^2 \\ &= \ell(u) - \alpha \int \nabla(e_N + \hat{e}^{pr} + \hat{e}^{du})^2 - a(e_N, e_N; \mu) + \alpha \int [\nabla e_N]^2 \ , \end{aligned} \tag{17}$$

and the proof follows from the uniform ellipticity of $a(.,.;\mu)$. □

Despite the fact that we have avoided to speak about any discretization so far, theorem 4.1 is already informative in the sense that in order to obtain s^-, the problem (16) to be solved, is parameter independent and simpler than the original one, provided that we have a good evaluation of the ellipticity constant. In section 6 we shall explain how to transform these constructions in a method that can be implemented. Before this we should explain how the previous estimator may help in designing a good choice for the elements of the reduced basis, providing a third alternative to the greedy algorithm presented in (7). Currently indeed, we have two alternative, either a random approach (that generally works not so badly) or select out of a large number of pre-computed solution $\{u_i\}_i$, the best sample from a SVD approach by reducing the matrix of scalar products (u_i, u_j). The former lacks of fiability, the latter is is a quite expensive approach and is mostly considered in a pre analysis framework as is explained in the next section. In order to reduce the cost of the off-line stage we can propose a greedy algorithm that combines the reduced approximation and the error evaluation :

- take a first parameter (randomly)
- use a (one dimensional) reduced basis approach over a set of parameter values (chosen randomly) and select, as a second parameter, the one for which the associated predicted error $s^+ - s^-$ is the largest.

this gives now a 2 dimensional reduced basis method.

- use this (2 dimensional) reduced basis approach over the same set of parameters and select, as a third parameter, the one for which the associated error is the largest.

this gives a 3 dimensional reduced basis method...

- and proceed...

Note that we then only compute accurately the solutions corresponding to the parameters that are selected this way.

The a posteriori approach that has been presented above relies on the uniform ellipticity of the bilinear form and the knowledge of the ellipticity constant. For more general problems, where only, nonuniform inf-sup conditions are valid (e.g. noncoercive Helmholtz acoustics problem which becomes singular as we approach resonance) smarter definitions should be considered. We refer to [18] for improved methods in this direction.

5. Some pragmatic considerations

Now that some basics on the reduced basis method have been presented, it is interesting to understand if the problem you have in mind is actually eligible to this type of approximation. We are thus going to propose some pragmatic arguments that may help in the preliminary verification. First of all, let us note that we have

illustrated the discretization on linear elliptic problems, of course this is just for the sake of simplicity, non linear problem [11, 19, 20] so as time dependent problems [7, 17] can be solved by these methods. Second, many things can be considered as a valid parameter: this can be the size of some simple geometric domain on which the solution is searched [16] , but it can be the shape itself [13] (the parameter in the former case is a multireal entity while in the latter it is a functional), the parameter can also be the time [7, 17] , or the position of some given singularities [2] ...

For all these choices, a fair regularity in the parameter is expected and wished so that the n-width goes fast to zero. An important remark has to be done here in order the size of the reduced basis be the smallest possible. Indeed, it may be smart to preprocess the precomputed solutions in order they look more similar. An example is given in [2] where quantum problem are considered; the solutions to these problems present some singularities at the position of the nuclei. If the position of the nuclei is the parameter we consider, it is useful to transform each solution in a reference configuration where the singularities/nuclei are at a unique place; the solutions are then much more comparable. Another example is given by the solution of the incompressible Stokes and Navier Stokes problem where the shape of the computational domain is the parameter; in order to be able to compare them properly, they have to be mapped on a unique (reference) domain. This is generally done through a simple change of variable. In case of the velocity, it is a vector field that is divergence free and a "standard" change of variable will (generally) not preserve this feature. The Piola transform (that actually corresponds to the classical change of variable over the potential function) allows to have the velocity fields transformed over the reference domain while preserving the divergence free condition as is demonstrated in [9]. These preprocessing steps allow to diminish the n-width of $\mathcal{S}(\mathcal{D})$ and it pays to be smart!!

In order to sustain the intuition on the potential of the reduced basis concept, a classical way is to use a SVD approach. Let us assume that we have a bunch of solutions $u_i = u(\mu_i)$, snapshots of the space $\mathcal{S}(\mathcal{D})$ of solutions to our problem. Out of these, the correlation matrix (u_i, u_j) which is symmetric can be reduced to its eigen-form, with positive eigenvectors that, ranked in decreasing order, go to zero. The high speed of convergence towards zero of the eigenvalues ranked in decreasing order will sustain the intuition that the reduced basis method will work. Indeed, the n-width is directly related to the size of the eigenvalues larger than the $n+1$th. The idea is that if the number of eigenvectors associated with the largest eigenvalues is small, then the method is viable. In order to sustain this, you can also consider, momentarily, the space X_N spanned by the eigenvectors associated with the N largest eigenvalues and analyze the norm of the difference between the snapshots in $\mathcal{S}(\mathcal{D})$ and their best fit in X_N. Note that we do not claim that this is a cheap constructive method: this procedure consists in a pre-analysis of the potential of the reduced basis method to approximate the problem you consider. If the norm of the error goes to zero sufficiently fast, you know that a Galerkin approach will provide the same order of convergence and the method is worth trying. We insist on the fact that this pre-analysis is not mendatory, it is only to

help in understanding what you should expect, "at best" from the reduced basis approximation. In particular the greedy approach presented in section 4 has to be preferred to the SVD approach that we discussed above for the determination of the elements that are to be incorporated in the reduced basis space, if you do not want to spend too much time during the off-line stage. Note also that the greedy approach provides solutions, that, when their number becomes large, become more and more linearly dependent (actually this is one of the aspects of the low n-width) and thus, for stability purposes it is important, through a Gramm-Schmidt process, to extract, from these solutions, orthonormal elements that will be the actual elements of the reduced basis: these will be named $(\zeta_i)_i$. This does not change the potential approximation properties of the reduced basis but improves, to a large extent, the stability of the implementation.

Finally, the preselection may be quite generous in the sense that you may be interested to select more than N basis functions, N being an evaluation of the dimension of the reduced basis for most problems. The reason for this comes from the conclusion of the a posteriori analysis that may tell you to increase the size of the reduced basis, suggesting you to work with $N+2$ (say) instead of N basis functions. This again is a feature of exponentially rapid convergence that lead to a large difference between the accuracy provided by X_N and X_{N+2} (say). It is time now to give some details on the implementation of the method.

6. Implementation issues.

We start by emphasizing tha any reduced basis method necessarily involves the implementation of a more "classical" approximation method. Indeed — except for very particular and uninteresting problems — the knowledge of the solutions, that we named u_i, is impossible without referring to a discretization method (e.g. of finite element, spectral type...). This is also the case for the ζ that are coming out from some shaping of the basis, e.g. Gram Schmidt, as explained earlier. This is the reason why reduced basis methods should not be considered as competitor to standard approximation methods but only as surrogates.

This implies, though, some difficulties since the elements of the *r*educed basis are only known through a preliminary *c*omputation basis, which, if we want the solution u_i to be well approximated, has to be *very* large. Knowing this, the rule of the game for the efficient implementation of any reduced basis method is to strictly prohibit any *online* reference to the extended basis. We allow *offline* pre-computations of the solutions (that involves the extended basis) and some *offline* cross contribution of these solutions (based on their expression with respect to the extended basis) but this is forbidden *online*. Following [16], we explain in the next subsection how this can be done.

6.1. Black box approach. The solution procedure involves the evaluation of the elements of the stiffness matrix $a(\zeta_i, \zeta_j; \mu), 1 \leq i, j \leq N$ that depends on the current parameter μ. This computation involves some derivatives and the

evaluation of integrals, that have to be performed and this may be *very* lengthy. It should be stated here that the implementation of the reduced type method has to be much faster than the solution procedure that was used to compute the reduced basis, much means many order of magnitude. The $\mathcal{O}(\dim X_N)^2$ entrees of the stiffness matrix have thus to be evaluated through some smart way.

Let us begin by the easy case that is named *affine parametric dependance* where the entries $a(\zeta_i, \zeta_j; \mu)$ appear to read

$$a(\zeta_i, \zeta_j; \mu) = \sum_p g_p(\mu) a_p(\zeta_n, \zeta_m) \ , \tag{18}$$

where the bilinear forms a_p are parameter independent. Many simple problems where the parameter are local constitutive coefficients or local zooming isotropic or non isotropic factors, enter in this framework. The expensive computation of the $a_{p,n,m} = a_p(\zeta_n, \zeta_m)$ can be done offline, once the reduced basis is constructed; these $a_{p,n,m}$ are stored and, for each new problem, the evaluation of the stiffness matrix is done, online, in $P \times N^2$ operations, and solved in $\mathcal{O}(dim X_N^3)$ operations. These figures are coherent with the rapid evaluation of the reduced basis method.

6.2. A posteriori implementation. Under the same affine dependance hypothesis on a, it is easy to explain how the a posteriori analysis can be implemented, resulting in a fast on-line solution procedure, provided some off-line computations are made. First of all the computation of $\tilde{\psi}_N$ can be implemented in the space $\tilde{X}_N = \text{Span}\{\xi_1, \ldots, \xi_N)\}$ exactly as above for the computation of u_N. Taking into account (18), together with the expressions obtained from the inversion of problem (5) and (12): $u_N = \sum_{i=1}^N \alpha_i \zeta_i$ and $\tilde{\psi}_N = \sum_{i=1}^N \tilde{\alpha}_i \xi_i$, we can write

$$\mathcal{R}^{pr}(v, \mu) = \sum_p \sum_i g_p(\mu) \alpha_i a_p(\zeta_i, v) - (f, v) \ ,$$

and

$$\mathcal{R}^{du}(v, \mu) = -\sum_p \sum_j g_p(\mu) \tilde{\alpha}_j a_p(v, \xi_j) - \ell(v) \ ,$$

hence by solving $n$umerically , off-line, each of the problems

$$2\alpha \int \nabla e^{pr,p,i} \nabla v = a_p(\zeta_i, v) \tag{19}$$

$$2\alpha \int \nabla e^{pr,0} \nabla v = (f, v) \tag{20}$$

$$2\alpha \int \nabla e^{du,p,j} \nabla v = a_p(v, \xi_j) \tag{21}$$

$$2\alpha \int \nabla e^{du,0} \nabla v = \ell(v) \ , \tag{22}$$

allows to write the numerical solutions of (16) as a linear combinaison of the elements previously computed (e.g. $\hat{e}^{pr} = \sum_p \sum_i g_p(\mu) \alpha_i e^{pr,p,i} - e^{pr,0}$) in $\mathcal{O}(PN)$ operations.

6.3. Magic points. The hypothesis of *affine parametric dependancy* is rather restrictive, and has to be generalized. In case of quadratic of cubic dependancy, this is quite straightforward but even for linear problems such as Laplace problem, when e.g. geometry is the parameter, this is rarely the case and another approach has to be designed. In order to get a better understanding of the method, let us first indicate that, when the geometry is the parameter, the solutions have to be mapped over a reference domain $\hat{\Omega}$. Let us assume that we want to compute $d(\zeta_i, \zeta_j; \Omega)$ where

$$d(u, v; \Omega) = \int_{\Omega} uv dA = \int_{\hat{\Omega}} uv J_{\Phi} d\hat{A} \ ,$$

where J_{Φ} is a Jacobian of the transformation that maps $\hat{\Omega}$ onto Ω. There is no reason in the general case that J_{Φ} will be affine so that the previous approach will not work. It is nevertheless likely that there exists a sequence of well chosen transformations Φ_1^*,..,Φ_M^*,... such that J_{Φ} may be well approximated by an expansion $J_{\Phi} \simeq \sum_{k=1}^{M} \beta_k J_{\Phi_k^*}$. An approximation of $d(\zeta_i, \zeta_j; \Omega)$ will then be given by

$$d(\zeta_i, \zeta_j; \Omega) \simeq \sum_{k=1}^{M} \beta_k \int_{\hat{\Omega}} \hat{\zeta}_i \hat{\zeta}_j J_{\Phi_k^*} d\hat{A} \ , \tag{23}$$

and again, the contributions $\int_{\hat{\Omega}} \hat{\zeta}_i \hat{\zeta}_j J_{\Phi_k^*} d\hat{A}$ will be pre-computed offline. We do not elaborate here on how the Φ_k^* are selected, and refer to [9], what we want to address is the evaluation of the coefficients $\beta_k = \beta_k(\Omega)$ in the approximation of J_{Φ} above. The idea is to use an interpolation procedure as is explained in [6]. Let $\boldsymbol{x}_1$ be the point where $|J_{\Phi_1^*}|$ achieves its maximum value. Assuming then that $\boldsymbol{x}_1$, ..., $\boldsymbol{x}_n$ have been defined, and are such that the $n \times n$ matrix with entries $J_{\Phi_k^*}(\boldsymbol{x}_\ell)$, $1 \leq k, \ell \leq n$ is invertible, we define $\boldsymbol{x}_{n+1}$ as being the point where $r_{n+1} = |J_{\Phi_{n+1}^*} - \sum_{k=1}^{n} \gamma_k J_{\Phi_k^*}|$ achieves it maximum value, here the scalar γ_k are defined so that r_{n+1} vanishes at any $(\boldsymbol{x}_\ell)$ for $\ell = 1, ..., n$. The definition of the points $\boldsymbol{x}_\ell$ is possible as long the Φ_ℓ are chosen such that the $J_{\Phi_\ell^*}$ are linearly independent (see [6]). The β_k are then evaluated also through the interpolation process

$$J_{\Phi}(\boldsymbol{x}_\ell) = \sum_{k=1}^{M} \beta_k J_{\Phi_k^*}(\boldsymbol{x}_\ell), \quad \forall 1 \leq \ell \leq M \ . \tag{24}$$

We have not much theory confirming the very good results that we obtain (which makes us call these interpolation point "magic"). An indicator that allows to be quite confident in the interpolation process is the fact that the Lebesgue constant attached to the previously built points is, in all example we have encountered, is rather limited.

Note that the same interpolation approach allows to compute the reconstructed errors with a compatible complexity as in the previous subsection.

The same magic point method has to be used also for the implementation of the reduced basis method for nonlinear problems. Actually, denoting by $z_i = \mathrm{NL}(u_i)$ the nonlinear expression involved in the problem, provided that the set

$Z_M = \text{Span}\{z_i, 1 \leq i \leq M\}$ has a small width, the interpolation process presented above allows both to determine a good interpolation set and a good associated interpolation nodes, we refer to ([6]) for more details on the implementation and to numerical results.

7. Some extensions

7.1. Eigenvalue problems. We end this paper by noticing that the reduced basis method can actually be found, at least in spirit, in many other approximations. There are indeed many numerical approaches that, in order to tackle a complex problem, use the knowledge of the solution of similar but simpler problems to facilitate the approximation. In this direction, the modal synthesis method provides a method to solve approximately eigenvalue problems on large structures based on the knowledge of the eigenvalues and eigenfunctions of the same problem on substructures. We refer e.g. to [4, 5] for more details on a high order implementation of these approaches.

Similarly, again, one of the approaches for the resolution of Hartree Fock problem in quantum chemistry is the L.C.A.O. method that consists in approximating the wave function of a molecule by linear combination of atomic orbitals that are nothing but solutions to the same problem on an atom, instead of a molecule. The atomic orbitals are indeed the approximations of Hydrogenoid functions (the contracted Gaussians have to be seen this way) that are the solutions of the electronic problem of one electron around a nucleus. This similarity is the guideline for the extension that is proposed in ([3, 2]).

At this level, it is also interesting to note that the reduced basis method, for an eigenvalue problem as the one encountered in the two previous cases, may be very appropriate since it can be proven that, letting u_i denote the set of all first P eigenvectors of an eigenvalue problem depending on a parameter μ, $u_i \equiv (e^1(\mu_i), .., e^p(\mu_i))$, then the approximation of this complete set of eigenvectors can be done with the same linear combinaison. More precisely it is possible to get an accurate approximation method based on

$$u(\mu) \simeq \sum_{i=1}^{P} \alpha_i u_i, \quad \forall j, \quad e^j(\mu) \simeq \sum_{i=1}^{P} \alpha_i e^j(\mu_i)$$

instead of

$$e^j(\mu) \simeq \sum_{i=1}^{P} \alpha_i^j e^j(\mu_i).$$

Again we refer to ([2]), for more details on this.

7.2. The reduced element method. In the reduced basis element method introduced in [13], we consider the geometry of the computational domain to be the generic parameter. The domain is decomposed into smaller blocks, all of them

can be viewed as the deformation of a few reference shapes. Associated with each reference shape are previously computed solutions (typically computed over different deformations of the reference shapes). The precomputed solutions are mapped from the reference shapes to the different blocks of the decomposed domain, and the solution on each block is found as a linear combination of the mapped precomputed solutions. The solutions on the different blocks are glued together using Lagrange multipliers.

To be more precise, we assume that the domain Ω where the computation should be performed can be written as the *non-overlapping* union of subdomains Ω^k:

$$\overline{\Omega} = \bigcup_{k=1}^{K} \overline{\Omega}^k, \quad \Omega^k \cap \Omega^\ell = \emptyset, \text{ for } k \neq \ell \ . \tag{25}$$

Next, we assume that each subdomain Ω^k is the deformation of the "reference" domain $\hat{\Omega}$ through a regular enough, and one to one, mapping. In an off-line stage, this reference geometry has been "filled up" with reference reduced basis solutions $\hat{u}_1, \hat{u}_2, .., \hat{u}_N$ to the problem that is under evaluation. Hence, together with this geometric decomposition, a functional decomposition is proposed since every Ω^k ; this allows us to define the finite dimensional space

$$Y_N = \{v \in L^2(\Omega), v_{|\Omega^k} = \sum_{i=1}^{N} \alpha_i^k \mathcal{F}_k^{-1}[\hat{u}_i]\} \ , \tag{26}$$

which is a set of uncoupled, element by element, discrete functions, where $\mathcal{F}_k$ allows to transform functions defined over $\hat{\Omega}$ into functions defined over Ω^k. This is generally not yet adequate for the approximation of the problem of interest since some glue at the interfaces $\gamma_{k,\ell}$ between two adjacent domains $\overline{\Omega}^k \cap \overline{\Omega}^\ell$ has to be added to the elements of Y_N, the glue depending on the type of equations we are interested to solve (it will be relaxed C^0–continuity condition for a Laplace operator, or more generally relaxed C^1–continuity condition for a fourth-order operator.

At this stage it should be noticed that, modulo an increase of complexity in the notations, there may exist not only one reference domain $\hat{\Omega}$ filled with its reduced basis functions but a few numbers so that the user can have more flexibilities in the design of the final global shape by assembling deformed basic shapes like a plumber would do for a central heating installation.

The reduced basis element method is then defined as a Galerkin approximation over the space X_N being defined from Y_N by imposing these relaxed continuity constraints. We refer to [9, 10] for more details on the implementation for for hierarchical fluid flow systems that can be decomposed into a series of pipes and bifurcations.

8. References

[1] Buffa, A., Maday, Y., Patera, A. T., Prud'homme, C., Turinici, G., in progress, 2006.

[2] Cancès, E., Le Bris, C., Maday, Y., Nguyen, N.C., Patera, A.T., Pau, G., Turinici, G., in progress (2006)

[3] Cancès, E., Le Bris, C., Maday, Y.,Turinici, G., Towards reduced basis approaches in ab initio electronic structure computations." Journal of Scientific Computing, 17(1-4):461–469, 2002.

[4] Charpentier, I., Maday, Y., Patera, A.T., Bounds evaluation for ouputs of eigenvalue problems approximated by the overlapping modal synthesis method, C. R. Acad. Sci. Paris, Srie I, 329, (1999), 909–914.

[5] Charpentier, I., de Vuyst, F. , Maday, Y., The overlapping component mode synthesis: The shifted eigenmodes strategy and the case of self-adjoint operators with discontinuous coefficients, Proceedings of the Ninth International Conference on Domain Decomposition Methods, P.E. Bjrstad, S. Magne, S. Espedal and D.E. Keyes Eds., (1998), 583–596.

[6] Grepl, M.A., Maday, Y., Nguyen, N.C., Patera, A.T., Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations, submitted to M2AN (2006).

[7] Grepl, M.A., Patera, A.T., A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations, M2AN, 39(1) (2005), 157–181.

[8] Kolmogoroff, A., Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse *Anals of Math.* **37** (1963), 107—110.

[9] Løvgren, A. E. , Maday, Y. , Rønquist, E. M., A reduced basis element method for the steady Stokes problem, to appear in M^2AN.

[10] Løvgren, A. E. , Maday, Y. , Rønquist, E. M., The reduced basis element method for fluid flows, to appear in Advances in Mathematical Fluid Mechanics (2006).

[11] Machiels, L. , Maday, M., Oliveira, I.B., Patera, A.T., Rovas, D.V., Output bounds for reduced-basis approximations of symmetric positive definite eigenvalue problems, CR Acad Sci Paris Series I 331, (2000), 153–158.

[12] Maday, Y., Patera, A. T., Turinici, G., A Priori Convergence Theory for Reduced-Basis Approximations of Single-Parameter Elliptic Partial Differential Equations, *J. Sci. Comput.* **17** (2002), no. 1-4, 437–446.

[13] Y. Maday and E. M. Rønquist — The reduced-basis element method: Application to a thermal fin problem. SIAM Journal on Scientific Computing, 2004.

[14] Pinkus, A., *n-Widths in Approximation Theory*, Springer-Verlag, Berlin,1985.

[15] Prud'homme, C., Contributions aux simulations temps-réel fiables et certains aspects du calcul scientifique, *Mémoire d'Habilitation à Diriger les Recherches*, Uiversité Pierre et Marie Curie, Paris 6, décembre 2005.

[16] Prud'homme, C., Rovas, D.V., Veroy, K ., Machiels, L., Maday, Y., Patera, A. T., Turinici, G., Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods, *J Fluids Engineering*, **124**, (2002) pp 70–80.

[17] Rovas, D. V. , Machiels, L. , Maday, Y., Reduced-basis output bound methods for parabolic problems IMA Journal of Numerical Analysis Advance Access published on March 6, (2006).

[18] Sen, S., Veroy, K., Huynh, D.B.P., Deparis, S. , Nguyen, N.C. , Patera, A.T. , "Natural norm"a posteriori error estimators for reduced basis approximations. Journal of Computational Physics, (2006)

[19] Veroy, K. , Prud'homme C., Patera, A.T., Reduced-basis approximation of the viscous Burgers equation: Rigorous a posteriori error bounds, CR Acad Sci Paris Series I 337, (2003), 619–624.

[20] Veroy, K. , Patera, A.T., Certified real-time solution of the parametrized steady incompressible Navier-Stokes equations: Rigorous reduced-basis a posteriori error bounds, Int. J. Numer. Meth. Fluids 47, (2005), 773–788.

Université Pierre et Marie Curie-Paris6, UMR 7598 Laboratoire Jacques-Louis Lions, B.C. 187, Paris, F-75005 France; and Division of Applied Maths, Brown University.

E-mail: maday@ann.jussieu.fr